Sox8 is sufficient to reprogram ectoderm into ear vesicles and associated neurons

Ailin Leticia Buzzi[1], Jingchen Chen[1,3,$], Alexandre Thiery[1,$], Julien Delile[2], Andrea Streit[1*]

[1] Centre for Craniofacial and Regenerative Biology, King's College London, London SE1 9RT, UK

[2] The Francis Crick Institute, London NW1 1AT, UK

[$] These authors contributed equally.

*Corresponding author: andrea.streit@kcl.ac.uk

[3] current address: Maijing Gene Diagnostics, Guangzhou 510030, China

The vertebrate inner ear arises from a pool of progenitors with the potential to give rise to all the sense organs and cranial ganglia of the head[1-6]. Here we explore the molecular mechanisms that control ear specification from these precursors. Using a multi-omics approach combined with loss-of-function experiments we identify a core transcriptional circuit that imparts ear identity, along with non-coding elements that integrate this information. This analysis places the transcription factor Sox8 at the top of the ear determination network. Introducing Sox8 into cranial ectoderm not only converts non-ear cells into ear progenitors, but also activates the cellular programmes for ear morphogenesis and neurogenesis. Thus, Sox8 emerges as a master regulator of ear identity and may be a key factor for sense organ cell reprogramming.

23 Master regulators have the unique ability to reprogram cells from other cell types. They activate a set
24 of transcription factors that in turn control the expression of proteins required for cell-specific functions.
25 The classic example is the transcription factor MyoD, which can convert fibroblasts into myogenic
26 cells[7]. From insects to vertebrates, the transcription factor Pax6 induces ectopic eyes when mis-
27 expressed in non-eye cells[8], however master regulators for other sense organs have not yet been
28 discovered. Here we set out to identify such factors for the organ of hearing and balance, the
29 vertebrate inner ear.

30 During development, the inner ear arises from the posterior border of the cranial neural plate.
31 Here, precursors for neural, neural crest, ear and sensory ganglia are initially intermingled, but
32 segregate over time[9,10]. FGF signalling from adjacent tissues induces otic-epibranchial progenitors
33 (OEPs), which subsequently converge to distinct territories: the otic and epibranchial placodes[11-15]
34 (Fig. 1a, e). While epibranchial cells produce sensory ganglia, the otic placode invaginates to form a
35 vesicle, which is then transformed into the inner ear, containing many specialised cell types and
36 associated neurons. Using transcriptional profiling, we recently defined the first gene regulatory
37 network for early ear development[16]. However, the key factors that impart ear character remain
38 unknown as do the non-coding regions that integrate the process. Here we use single cell RNA-
39 sequencing (scRNAseq) and pseudo-time analysis to predict transcriptional changes as progenitors
40 undergo ear commitment. Using epigenomic profiling we also determine ear enhancers and their
41 upstream regulators. To ensure transparency and reproducibility, we have developed a fully
42 integrated workflow of the data analysis in Nextflow (available at https://alexthiery.github.io/otic-
43 reprogramming/). Together with functional experiments, we identify a minimal transcriptional circuit
44 that defines ear identity comprising Sox8, Pax2 and Lmx1a, with Sox8 at the top of the hierarchy.
45 Introduction of Sox8 alone into cranial ectoderm is sufficient to trigger the ear programme, to initiate
46 ear morphogenesis by forming ear vesicles and to activate neurogenesis. Thus, using a multi-omics
47 approach we have uncovered Sox8 as a master regulator of ear fate, which may in the future allow
48 reprograming of specific sensory cell types.
49

50 **Dynamic changes in gene expression characterise the transition from progenitor to otic**
51 **commitment**

52 To unravel the genetic hierarchy that controls how otic and epibranchial cells diverge, we first
53 characterised the transcriptional profile of cells committed to each fate[2] (Fig1. a-d, Extended Data
54 Fig. 1a, b). To label each cell population we used two enhancers driving EGFP, the novel otic
55 Lmx1aE1 enhancer (see below) and the epibranchial enhancer Sox3U3[17]. Reporter constructs were
56 electroporated into the OEP territory of somite stage (ss) 3-5 embryos. At ss18-21, EGFP$^+$ cells were
57 isolated by fluorescence-activated cell sorting (FACS) and processed for RNA-seq. Differential
58 expression analysis identifies 103 and 319 genes upregulated in Lmx1aE1-EGFP and Sox3U3-EGFP
59 expressing cells, respectively (log2FC > 1.5; adjusted p-value < 0.05; Fig. 1d; Extended Data Fig. 1c,
60 d and Supplementary Table1). This allows us to define the transcriptional states of definitive otic and
61 epibranchial cells.

62 Next, we investigated the transcriptional changes which take place as OEPs transition from a
63 common progenitor population to definitive otic and epibranchial cells. *Pax2* is expressed throughout
64 this time-window in both cell populations (Extended Data Fig.1e, f)[2,9]. Using a Pax2E1-EGFP reporter,
65 we isolated single cells by FACS from consecutive stages of ear specification: OEP (ss8-9), early-
66 placode (ss11-12) and late-placode stage (ss14-15) and processed them for single-cell-RNA-seq
67 (scRNAseq; Fig. 1e, f). To characterise cellular diversity, we first looked at groups of genes co-
68 expressed across the dataset (gene modules). These gene modules were identified in an unbiased
69 manner through hierarchical clustering of gene-gene Spearman correlation coefficient[18].Gene
70 modules of interest were selected based on the presence of well-characterized makers for placodal,
71 neural and neural-crest cells including the new otic or epibranchial genes identified above (Fig.1d).
72 Initial cell clustering defines five major clusters (Extended Data Fig. 2a) to which we assigned
73 identities using known markers (Fig. 1g-h, Extended Data Fig. 2c-f). Clusters C1 and C2 represent
74 *Pax2*$^+$/*Six1*$^+$ cells expressing high levels of OEP and placodal makers (Fig. 1h, i; Extended Data Fig.
75 2c). Surprisingly, we also find two clusters with relatively few *Pax2*$^+$ cells: one containing neural-like
76 cells (C4; *Sox21*$^+$) and another containing neural crest-like cells (C5; *Pax7*$^+$) (Extended Data Fig. 2a,
77 d, e). Since OEPs are mixed with neural and neural crest cells at early stages[9], this observation

78 suggests that while cells in clusters C4/5 initially activate Pax2E1-EGFP they subsequently
79 downregulate enhancer activity and *Pax2* expression.

80 To investigate the transcriptional dynamics accompanying otic and epibranchial fate
81 decisions, we subset the placodal clusters (C1/2 in Fig. 1; Extended Data Fig. 2a). Re-clustering these
82 cells using gene modules containing otic and epibranchial genes we obtained five placodal clusters
83 (PC1-5; Fig. 1m, j, k; Supplementary Table 2). PC1 largely consists of only OEPs (ss8-9 cells), while
84 the other clusters contain cells from both early and late placode stages (Fig.1m, j, k). Indeed, PC1 is
85 characterised by the expression of OEP genes (GM1-3; Fig. 1m), as well as sharing genes with the
86 otic module GM5/6 and the epibranchial module GM7-9. In contrast, PC2/3 and PC4/5 are
87 transcriptionally distinct from each other with profiles akin to otic (GM5/6) and epibranchial (GM7-9)
88 cells, respectively. To explore the relationship between different cell clusters we organised cells along
89 pseudo-time using Monocle2[19]. This analysis predicts that OEPs gradually split into one otic and one
90 epibranchial branch, each composed of early and late placodal cells (Fig.1n, o). Calculating RNA-
91 velocity independently[20] and embedding the corresponding vector field onto the Monocle2 trajectory
92 validates the directionality of predicted cell state transitions (Fig.1p).

93 To explore dynamic changes of gene expression accompanying these inferred trajectories,
94 we used Branch-Expression-Analysis-Modelling (BEAM)[19]. This identified groups of transcription
95 factors expressed in OEPs prior to the branching point, which subsequently segregate into either the
96 otic (e.g. *Sox8, Lmx1a*) or the epibranchial (e.g. *Foxi3, Tfap2e*) branch (Fig. 2a, Extended Data Fig.
97 3a). To quantify the changes in co-expression of otic and epibranchial genes, we assessed the
98 proportion of co-expressing cells before and after the branching point. A two-tailed Wilcoxon rank sum
99 test reveals significantly more cells co-expressing otic and epibranchial markers in OEPs than in
100 epibranchial ($W=214$, $p=0.0013$) and otic cells ($W=235$, $p<0.0001$) after the branching point (Fig. 2b,
101 c). Quantification of gene expression in the monocle trajectories and by *in situ* hybridization chain
102 reaction (HCR)[21] confirms that otic (*Sox8, Lmx1a*) and epibranchial (*Foxi3, Tfap2e*) transcripts overlap
103 at ss8-9, and their expression resolves as both placodes are firmly established (Fig. 2d-i, Extended
104 Data Fig. 3b-f).

105 Together, these results identify groups of transcription factors whose expression changes over
106 time as otic and epibranchial precursors segregate. Some of these may play a key role in imparting
107 ear identity.
108
109 **Epigenomic profiling uncovers new regulatory elements and motifs in ear precursors**
110 Transcription factors controlling cell fate choice regulate cell-specific downstream targets by
111 interacting with tissue specific cis-regulatory enhancer elements. Active enhancers are regions of
112 open chromatin flanked by nucleosomes enriched for histone 3 lysine-27 acetylation (H3K27ac), while
113 actively transcribed genes are marked by H3K4me3[22-25]. We therefore profiled ss8-9 OEPs by
114 ChIPseq for H3K27ac, H3K4me3 and the repressive mark H3K27me3 and determined chromatin
115 accessibility by ATACseq. Overlapping H3K27ac and ATACseq data identifies 10969 regions with
116 depleted H3K27me3 marks; average profiles show bimodal H3K27ac read distribution surrounding
117 ATACseq peaks (Extended Data Fig. 4a). Of these just over 70% are intergenic or intronic
118 representing putative enhancers (8316), while the remaining are close to transcription start sites (TSS;
119 Extended Data Fig. 3b). We associated each putative enhancer to the nearest TSS of protein coding
120 genes; GO term analysis of the corresponding genes returns MAP-kinase, Wnt and Notch signalling
121 known to mediate ear induction, development and neurogenesis (Extended Data Fig. 4c)[13,15]. We
122 selected putative enhancers in the vicinity of ear-enriched genes to assess their activity *in vivo*,
123 generated EGFP reporters and co-electroporated them with ubiquitously expressed mCherry into
124 head-fold-stage chick embryos. RT-PCR-based assays[26] (not shown) and fluorescence microscopy
125 confirm enhancer activity in ear progenitors and otic placodes (Fig. 3a, b; Extended Data Fig. 5-7).
126 Next, we performed motif enrichment analysis of all ear enhancers to identify upstream regulators
127 that may act as otic determinants. This reveals an over-representation of binding sites for Sox, TEAD
128 and Six family members and for Tfap2a (Fig. 3c; Extended Data Fig. 4d). Of these, Tfap2a and Six
129 proteins have previously been implicated in cranial placode development[27-31], confirming that this
130 strategy can identify relevant regulatory factors.
131 We also exploited the idea that cell identity genes may be regulated by super-enhancers
132 characterised by broad H3K27ac, while their gene bodies harbour broad H3K4me3[32-34]. Focusing on

133 OEP factors segregating to the ear lineage (Fig. 2a) that are also prominent in the otic GRN[16], we find
134 that the Sox8 locus is marked with H3K4me3, while enhancers close to Lmx1a, Zbtb16 and Sox13
135 show broad H3K27ac marks (Fig. 3a, Extended data 4e, 6a, 7a). Together, this analysis points to
136 Lmx1a, Zbtb16 and members of the Sox transcription factor family as potential key regulators of ear
137 identity.
138

139 **Sox8, Lmx1a and Pax2 are core components of the ear determination network**

140 Our transcriptome analysis identifies several Sox genes enriched in otic progenitors. Of these, *Sox3*
141 and *Sox13* are expressed prior to ear specification, while *Sox9* and *-10* are activated later[16]. These
142 genes are therefore unlikely to initiate the otic programme. In contrast, the SoxE group factor *Sox8* is
143 highly enriched in OEPs prior to segregating to otic cells (Fig. 3d). *In situ* hybridisation reveals that
144 *Sox8* begins to be expressed at 3ss, followed shortly thereafter by *Pax2*, *Zbtb16*, and *Lmx1a*, while
145 *Foxg1*, *Soho1* and *Sox10* are activated later (Fig. 3d, Extended Data Fig. 5a).

146 To explore the regulatory interactions between the earliest OEP transcription factors, we
147 systematically knocked down each one in turn and assayed the expression of all others as well as
148 that of *Foxg1* and *Soho1* as a readout for otic identity using *in situ* hybridisation and RT-qPCR (Fig.
149 2e; Extended Data Fig. 5b, c). Control or antisense oligonucleotides targeting *Sox8, Pax2, Zbtb16* or
150 *Lmx1a* were electroporated into future OEPs of head-fold-stage chick embryos and gene expression
151 analysed at OEP stage (ss8-9). We find that *Sox8* is necessary for the expression of all assayed ear
152 transcription factors. *Pax2* is required for the expression of *Zbtb16* and *Lmx1a*, which in turn is
153 necessary for *Pax2,* suggesting that they act in a positive feedback loop. All gene expression changes
154 can be rescued by co-electroporation of the appropriate full-length constructs (Extended Data Fig. 9).
155 These experiments put Sox8 at the top of the ear determination network forming a regulatory circuit
156 with Pax2 and Lmx1a.
157

158 **Sox8 induces ectopic otic vesicles and vesicle-derived neurons**

159 If Sox8, Pax2 and Lmx1a indeed form a minimal circuit driving otic specification, they should be able
160 to convert non-ear cells into cells with ear identity. We tested this hypothesis by electroporating a

161 combination of Sox8/Pax2/Lmx1a-mCherry-tagged constructs into head-fold-stage ectoderm not
162 destined to contribute to the ear. After 72hrs we observe many otic vesicles scattered across the head
163 ectoderm. Some of these contain neurofilament-positive neurons, with processes projecting out from
164 the ectopic vesicle (Fig. 4a).

165 We next asked whether Sox8 alone can initiate the otic programme using the Lmx1aE1-EGFP
166 enhancer and otic markers as a readout. We find that misexpression of Sox8-mCherry not only
167 activates Lmx1aE1-EGFP in ectopic vesicles, but also the expression of *Pax2* and *Lmx1a*, and the
168 ear marker *Soho-1* (Fig. 4b-d; Extended Data Fig. 10a). To assess to what extent Sox8 can confer
169 ear identity, we isolated double-positive Sox8-mCherry/Lmx1a-EGFP cells from ss11-12 by FACS
170 and compared their transcriptome with control ectoderm labelled with constitutive mCherry/EGFP.
171 Differential expression analysis shows upregulation of 399 transcripts in comparison to controls, while
172 112 genes are downregulated (log2FC > 1.5; adjusted p-value < 0.05) (Fig. 4e, Extended Data Fig.
173 9b-d, Supplementary Table 3). Seventeen of the twenty seven upregulated transcription factors are
174 known to be expressed in the otic placode or vesicle, while for the remaining ten no expression data
175 are available (Supplementary Table 3). In contrast, among the downregulated genes are typical neural
176 crest and forebrain transcription factors. The first cell types specified during ear formation are neurons
177 of the cochlear-vestibular ganglion, we therefore asked whether Sox8 can also trigger neurogenesis.
178 We find that Sox8-induced ectopic vesicles are associated with neurofilament-positive neurites
179 generated from Sox8 expressing cells themselves (Fig. 4f, g). Together, these results show that Sox8
180 can activate the transcriptional program for ear identity in non-ear cells by direct lineage reprograming
181 *in vivo*.

182

183 **Conclusion**

184 We have established an integrated workflow for multi-omics data analysis to ensure accessibility and
185 reproducibility. Using transcriptional profiling at single cell level, enhancer discovery and analysis,
186 together with functional experiments, we have identified the critical components of the ear
187 determination network: Sox8, Pax2 and Lmx1a (Fig. 4h). Sox8 lies at the top of the transcriptional
188 hierarchy, and alone is sufficient to impart ear identity to cells otherwise destined to form head

189 epidermis, other sense organs or cranial ganglia. Previous findings have implicated Pax2 in otic
190 placode development and patterning[35-38] and Sox8 in regulating ear-specific enhancers[39-41]. Sox
191 proteins cooperate with a variety of transcription factors to exert their cell type specific function[42]. It is
192 therefore tempting to speculate that, in analogy to the eye, where Sox2 cooperates with Pax6 to
193 regulate lens-specific transcription[43], in the ear Sox8 might partner with Pax2.

194 Our results also show that OEPs initially express competing transcriptional programmes that
195 resolve over time as otic and epibranchial cell states are established. We capture previously unknown
196 gene modules that accompany this process as well as regulatory regions associated with otic-
197 epibranchial specification. In turn, this information is critical to unravel the underlying gene regulatory
198 networks and identify the transcription factor codes that determine cell identity in the cranial sensory
199 nervous system. In the long term, this will enhance our ability to engineer specific sensory cell types
200 for basic research and regenerative purposes.

201

202 **References**

203 1 Jacobson, A. G. The determination and positioning of the nose, lens, and ear. III. Effects of
204 reversing the antero-posterior axis of epidermis, neural plate and neural fold. *J Exp Zool*
205 **154**, 293-303 (1963).

206 2 Groves, A. K. & Bronner-Fraser, M. Competence, specification and commitment in otic
207 placode induction. *Development* **127**, 3489-3499 (2000).

208 3 Baker, C. V., Stark, M. R., Marcelle, C. & Bronner-Fraser, M. Competence, specification and
209 induction of Pax-3 in the trigeminal placode. *Development* **126**, 147-156 (1999).

210 4 Bhattacharyya, S. & Bronner-Fraser, M. Competence, specification and commitment to an
211 olfactory placode fate. *Development* **135**, 4165-4177 (2008).

212 5 Schlosser, G. Early embryonic specification of vertebrate cranial placodes. *Wiley Interdiscip*
213 *Rev Dev Biol* **3**, 349-363, doi:10.1002/wdev.142 (2014).

214 6 Thiery, A., Buzzi, A. L. & Streit, A. Cell fate decisions during the development of the
215 peripheral nervous system in the vertebrate head. *Curr Top Dev Biol* **139**, 127-167,
216 doi:10.1016/bs.ctdb.2020.04.002 (2020).

217 7 Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts
218 fibroblasts to myoblasts. *Cell* **51**, 987-1000, doi:10.1016/0092-8674(87)90585-x (1987).
219 8 Gehring, W. J. The master control gene for morphogenesis and evolution of the eye. *Genes*
220 *Cells* **1**, 11-15. (1996).
221 9 Streit, A. Extensive cell movements accompany formation of the otic placode. *Dev Biol* **249**,
222 237-254 (2002).
223 10 Pieper, M., Eagleson, G. W., Wosniok, W. & Schlosser, G. Origin and segregation of cranial
224 placodes in Xenopus laevis. *Dev Biol* **360**, 257-275, doi:S0012-1606(11)01268-1 [pii]
225 10.1016/j.ydbio.2011.09.024 (2011).
226 11 Ladher, R. K., Anakwe, K. U., Gurney, A. L., Schoenwolf, G. C. & Francis-West, P. H.
227 Identification of synergistic signals initiating inner ear development. *Science* **290**, 1965-1968.
228 (2000).
229 12 Wright, T. J. & Mansour, S. L. Fgf3 and Fgf10 are required for mouse otic placode induction.
230 *Development* **130**, 3379-3390 (2003).
231 13 Ohyama, T., Groves, A. K. & Martin, K. The first steps towards hearing: mechanisms of otic
232 placode induction. *Int J Dev Biol* **51**, 463-472 (2007).
233 14 Ladher, R. K., O'Neill, P. & Begbie, J. From shared lineage to distinct functions: the
234 development of the inner ear and epibranchial placodes. *Development* **137**, 1777-1785,
235 doi:137/11/1777 [pii] 10.1242/dev.040055 (2010).
236 15 Chen, J. & Streit, A. Induction of the inner ear: stepwise specification of otic fate from
237 multipotent progenitors. *Hear Res* **297**, 3-12, doi:10.1016/j.heares.2012.11.018 (2013).
238 16 Chen, J. *et al.* A systems-level approach reveals new gene regulatory modules in the
239 developing ear. *Development* **144**, 1531-1543, doi:10.1242/dev.148494 (2017).
240 17 Nishimura, N. *et al.* A Systematic Survey and Characterization of Enhancers that Regulate
241 Sox3 in Neuro-Sensory Development in Comparison with Sox2 Enhancers. *Biology (Basel)*
242 **1**, 714-735, doi:10.3390/biology1030714 (2012).

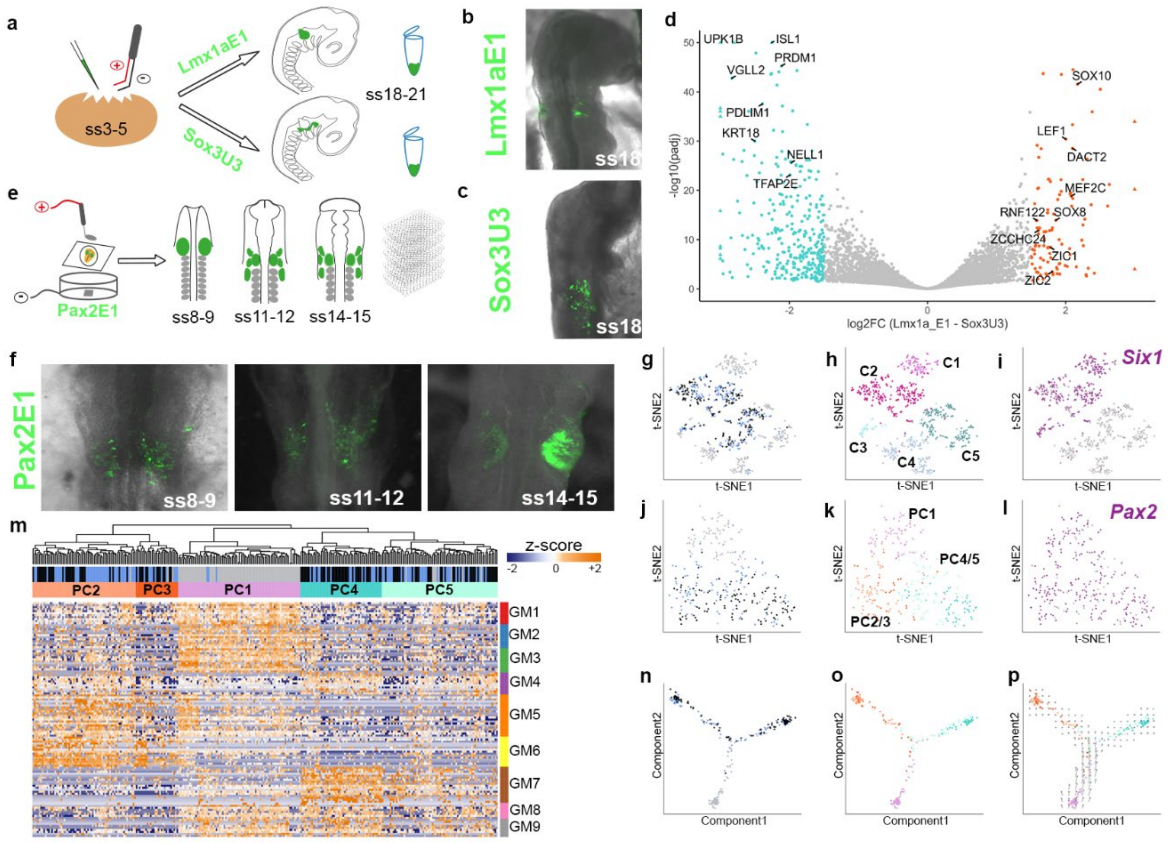
243 18 Delile, J. *et al.* Single cell transcriptomics reveals spatial and temporal dynamics of gene
244 expression in the developing mouse spinal cord. *Development* **146**, doi:10.1242/dev.173807
245 (2019).
246 19 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by
247 pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859
248 (2014).
249 20 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-
250 018-0414-6 (2018).
251 21 Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed,
252 quantitative, sensitive, versatile, robust. *Development* **145**, doi:10.1242/dev.165753 (2018).
253 22 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and
254 predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936,
255 doi:10.1073/pnas.1016071107 (2010).
256 23 Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila
257 melanogaster. *Nature* **471**, 480-485, doi:10.1038/nature09725 (2011).
258 24 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental
259 enhancers in humans. *Nature* **470**, 279-283, doi:nature09692 [pii] 10.1038/nature09692
260 (2011).
261 25 Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple
262 classes of enhancers with distinct cellular functions. *Genome Res* **21**, 1273-1283,
263 doi:10.1101/gr.122382.111 (2011).
264 26 Chen, J. & Streit, A. A medium-scale assay for enhancer validation in amniotes. *Dev Dyn*
265 **244**, 1291-1299, doi:10.1002/dvdy.24306 (2015).
266 27 Brugmann, S. A., Pandur, P. D., Kenyon, K. L., Pignoni, F. & Moody, S. A. Six1 promotes a
267 placodal fate within the lateral neurogenic ectoderm by functioning as both a transcriptional
268 activator and repressor. *Development* **131**, 5871-5881 (2004).

269 28 Ozaki, H. *et al.* Six1 controls patterning of the mouse otic vesicle. *Development* **131**, 551-
270 562 (2004).
271 29 Zheng, W. *et al.* The role of Six1 in mammalian auditory system development. *Development*
272 **130**, 3989-4000 (2003).
273 30 Maharana, S. K. & Schlosser, G. A gene regulatory network underlying the formation of pre-
274 placodal ectoderm in Xenopus laevis. *BMC Biol* **16**, 79, doi:10.1186/s12915-018-0540-5
275 (2018).
276 31 Kwon, H. J., Bhat, N., Sweet, E. M., Cornell, R. A. & Riley, B. B. Identification of early
277 requirements for preplacodal ectoderm and sensory organ development. *PLoS Genet* **6**,
278 doi:10.1371/journal.pgen.1001133 (2010).
279 32 Benayoun, B. A. *et al.* H3K4me3 breadth is linked to cell identity and transcriptional
280 consistency. *Cell* **158**, 673-688, doi:10.1016/j.cell.2014.06.027 (2014).
281 33 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-
282 947, doi:10.1016/j.cell.2013.09.053 (2013).
283 34 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at
284 key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
285 35 Torres, M., Gomez-Pardo, E. & Gruss, P. Pax2 contributes to inner ear patterning and optic
286 nerve trajectory. *Development* **122**, 3381-3391 (1996).
287 36 Burton, Q., Cole, L. K., Mulheisen, M., Chang, W. & Wu, D. K. The role of Pax2 in mouse
288 inner ear development. *Dev Biol* **272**, 161-175 (2004).
289 37 Christophorou, N. A., Mende, M., Lleras-Forero, L., Grocott, T. & Streit, A. Pax2 coordinates
290 epithelial morphogenesis and cell fate in the inner ear. *Dev Biol* **345**, 180-190, doi:S0012-
291 1606(10)00911-5 [pii] 10.1016/j.ydbio.2010.07.007 (2010).
292 38 Freter, S. *et al.* Pax2 modulates proliferation during specification of the otic and epibranchial
293 placodes. *Dev Dyn* **241**, 1716-1728, doi:10.1002/dvdy.23856 (2012).

294 39 Betancur, P., Sauka-Spengler, T. & Bronner, M. A Sox10 enhancer element common to the
295 otic placode and neural crest is activated by tissue-specific paralogs. *Development* **138**,
296 3689-3698, doi:dev.057836 [pii] 10.1242/dev.057836 (2011).

297 40 Murko, C. & Bronner, M. E. Tissue specific regulation of the chick Sox10E1 enhancer by
298 different Sox family members. *Dev Biol* **422**, 47-57, doi:10.1016/j.ydbio.2016.12.004 (2017).
299 41 Okamoto, Y. *et al.* Cooperation of Sall4 and Sox8 transcription factors in the regulation of the
300 chicken Sox3 gene during otic placode development. *Dev Growth Differ* **60**, 133-145,
301 doi:10.1111/dgd.12427 (2018).
302 42 Kondoh, H. & Kamachi, Y. SOX-partner code for cell specification: Regulatory target
303 selection and underlying molecular mechanisms. *The International Journal of Biochemistry &*
304 *Cell Biology* **42**, 391-399 (2010).
305 43 Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. Pax6 and SOX2 form a
306 co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*
307 **15**, 1272-1286 (2001).

308 **Figures and Figure Legends**



322 **Figure 1. Transcriptomic characterisation of ear development. a-c**, *In ovo* electroporation (**a**) was
323 used to label and collect otic (**b**; Lmx1aE1-EGFP^{+}) and epibranchial (**c;** Sox3U3-EGFP^{+}) cells for bulk
324 RNAseq. **d**, Volcano plot showing genes differentially expressed (absolute log2 fold-change >1.5 and
325 adjusted p-value < 0.05) in otic (orange) and epibranchial cells (green). **e**, **f**, Cells were collected for
326 scRNAseq after *ex ovo* electroporation of the Pax2E1-EGFP enhancer, which is active in OEPs, otic
327 and epibranchial placodes (**f**); cells were collected at the stages indicated. **g-l**, tSNE representation
328 of unsupervised hierarchical clustering of all cells (**g-i**) and of the placodal cell subset (**j-l**); cells were
329 colour-coded according to stage collected (**g**, **j**; OEP grey, ss11-12 blue, ss14-15 black), clusters (**h**,
330 **k**) and placodal markers *Six1* (**i**) and *Pax2* (**l**) expression. **m**, Heatmap showing partitioning of the
331 placodal subset (C1, C2 in h) into five clusters (PC1-5) based on gene modules (GM). Expression
332 profiles reveal that PC1 largely contains OEP-like cells, while PC2/ and PC4/5 are composed of otic-
333 like and epibranchial-like cells. **n-o**, Pseudo-time ordering using Monocle2 shows trajectories
334 between stages (**n**) and clusters (**o**). Note: otic- and epibranchial-like cells segregate into two
335 branches. **p**, RNAvelocity vector field verifies the directional trajectories predicted by Monocle2.

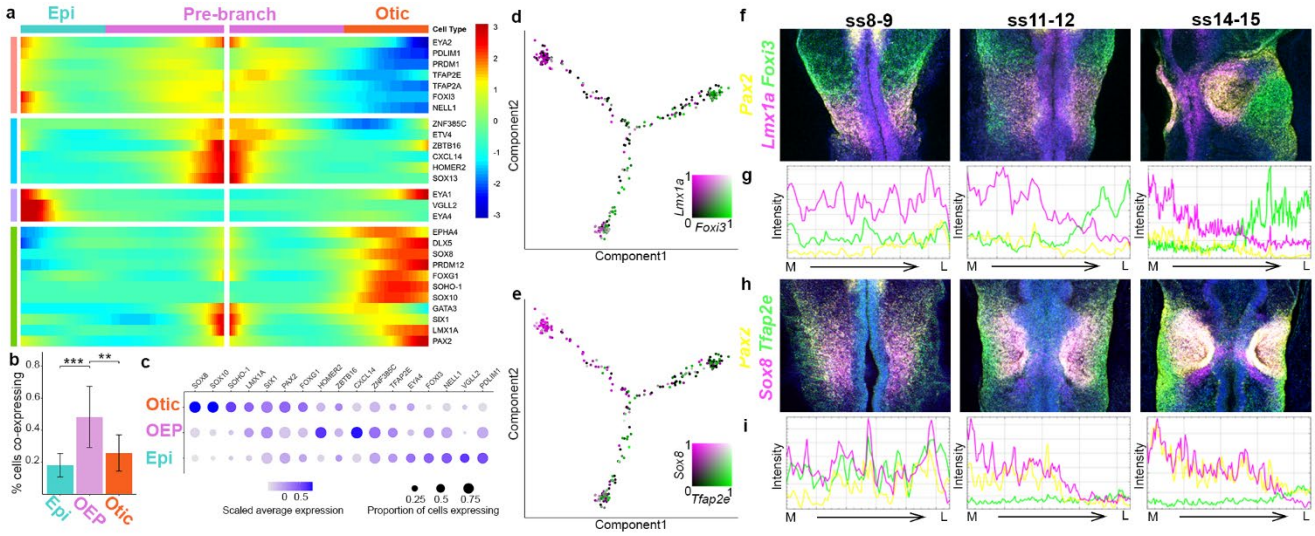


Figure 2. Dynamic gene expression during OEP segregation into otic and epibranchial lineages. a, BEAM identifies genes regulated in a branch-dependant manner. **b**, Histogram showing the proportion of cells co-expressing genes that are expressed before the branching point, but segregate to the otic or epibranchial branch. Significantly more cells co-express such genes before the branching point than thereafter. Error bars indicate +/- 1 standard deviation. ** p-value < 0.01; *** p-value < 0.001. **c**, Dot plot for OEP, otic and epibranchial markers based on scRNAseq data. Expression level is indicated by colour intensity and gene expression frequency by dot size. **d**, **e**, A proportion of cells co-expresses otic (*Sox8*, *Lmx1a*) and epibranchial (*Foxi3*, *Ap2e*) markers prior to the branching point. **f-i**, *In situ* HCR (**f**, **h**) and intensity measurements (**g**, **i**) along the medial (M) to lateral (L) axis of the OEP-domain confirms gradual segregation of OEP gene expression as distinct otic and epibranchial cell populations emerge.

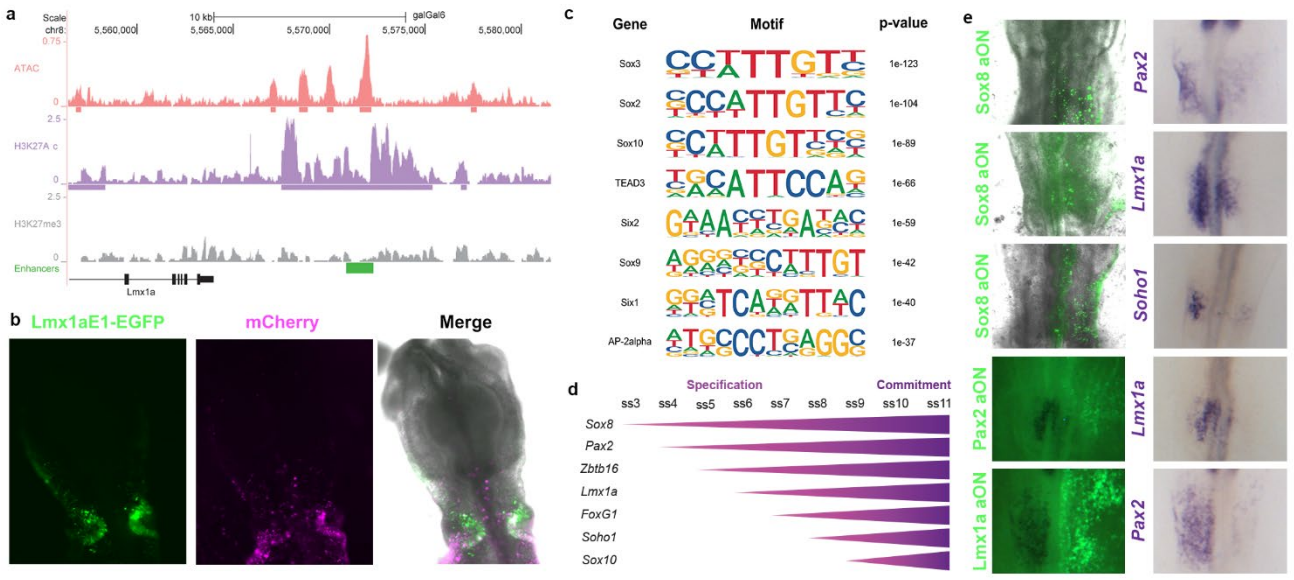


Figure 3. Chromatin accessibility profile of OEPs identifies putative regulatory regions. a, Genome browser view of ATAC, H3K27ac and H3K27me3 profiles in OEPs at the Lmx1a locus. Cloned putative enhancer is shown in green. **b**, Coelectroporation of Lmx1aE1-EGFP reporter (green) and constitutive mCherry (magenta) reveals *in vivo* enhancer activity in the otic placode. **c**, Motif enrichment analysis of all identified enhancers. **d**, Diagram showing the activation of expression of potential regulators of otic identity during specification and commitment stages. **e**, Knock-down of selected transcription factors using fluorescein labelled antisense oligonucleotides (aON; green) lead to downregulation of otic markers as identified by *in situ* hybridisation (blue).

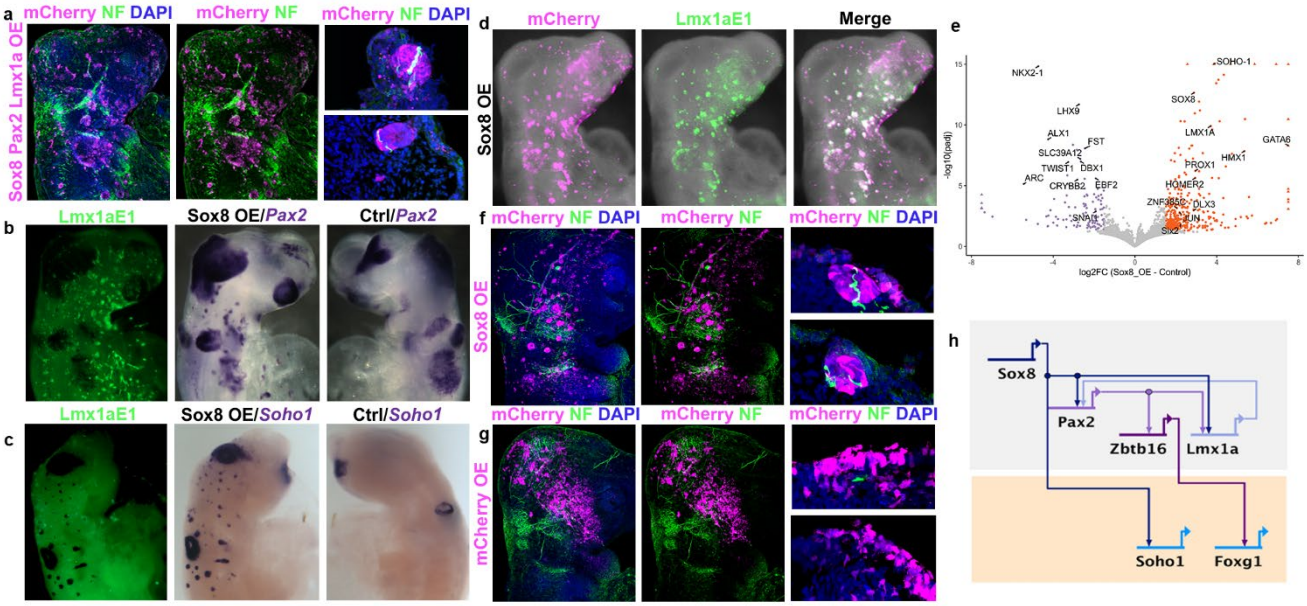


Figure 4. Direct reprogramming of ectoderm by Sox8. a, Co-electroporation of mCherry-tagged vectors containing the full-length sequence of Sox8/Pax2/Lmx1a (magenta) generate ectopic vesicles across the cranial ectoderm after 72hrs culture. Neurofilament antibody staining (NF; green) shows axonal projections from the vesicle. **b-d**, Overexpression of Sox8 (Sox8 OE) alone activates Lmx1aE1-EGFP (green) and otic marker expression on the electroporated, but not on the control side (Ctrl). **e**, RNAseq of Sox8 and control electroporated cranial ectoderm; volcano plot shows enrichment of otic genes in the Sox8 OE (absolute log2 fold-change >1.5 and adjusted p-value < 0.05). **f**, Sox8-mCherry overexpression (magenta) alone generates ectopic otic vesicles with neuronal projections (green), while controls do not (**g**). **h**, BioTapestry model showing the minimal transcriptional circuit for otic specification with Sox8 at the top of the hierarchy.

372 **Methods**

373 **Expression and enhancer constructs**

374 Putative enhancers were amplified from chick genomic DNA and cloned into pTK-EGFP reporter
375 vectors after digestion with XcmI[26]. To generate expression constructs total RNA was isolated from
376 HH8-12 chick embryos with RNAqueous-Micro Total RNA Isolation Kit (Thermo Fisher Scientific),
377 reverse transcribed using Superscript II reverse transcriptase (Thermo Fisher Scientific, 180644-014)
378 and oligo-dT primer. Specific primers were used to amplify the full-length coding sequence of Sox8,
379 Pax2, Lmx1a and Zbtb16 and PCR products were cloned into pCAB-IRES-mCherry. All sequences
380 were verified by Sanger sequencing.

381

382 **Chick embryos, electroporation and culture**

383 Fertilised hens' eggs (Stewart, Norfolk UK) were incubated at 38°C and staged according to
384 Hamburger and Hamilton (HH)[44]. All experiments were performed on embryos younger than 12 days,
385 thus were not regulated by the Animals Scientific Procedures Act 1986.

386 OEPs for bulk RNAseq were labelled using *in ovo* electroporation[45]; eggs were incubated until
387 the 3-6 somite stage (ss), pTK-Lmx1aE1-EGFP or pTK-Sox3U3-EGFP plasmids (1µg/µl) were
388 injected targeting the OEP territory. Electroporation using Ovodyne electroporator (TSS20, Intracel)
389 was performed with five 50ms pulses of 8V at 100ms intervals. After incubation until ss18-21, embryos
390 were collected in PBS and processed for bulk RNAseq.

391 For *ex ovo* culture, embryos were harvested using filter paper rings[46] and cultured on egg albumen;
392 for long-term culture (72hrs) the 'modified Cornish pasty' method[47] was used. *Ex ovo* electroporation
393 was performed to collect cells for scRNAseq, for knock-down, rescue and overexpression
394 experiments. The posterior placodal region or the cranial surface ectoderm of HH6 embryos was
395 targeted for electroporation by injecting plasmid DNA at 1µg/ul (pTK-Pax2E1-EGFP to label OEPs;
396 pCES-Sox8-mCherry + pTK-Lmx1aE1-EGFP, pCES-Sox8-mCherry + pCES-Lmx1a-mCherry +
397 pCES-Pax2-mCherry; pCES-mCherry + pCES-EGFP for misexpression), control or antisense-
398 oligonucleotides (aON; GeneTools; 0.75mM with 0.3µg/µl carrier DNA), or a combination of aON and
399 expression construct (0.75mM aON + 0.75µg/µl plasmid). Oligonucleotides used were Sox8: 5'-

400 CTCCTCGGTCATGTTGAGCATTTGG-3'[39], Pax2: 5'-GGTCTGCCTTGCAGTGCATATCCAT-3'[37],
401 Lmx1a: 5'-CCTCCATCTTCAAGCCGTCCAGCAT-3'[16], Zbtb16: 5'-
402 GTCAAATCCATAGCACTCCCGAGGT-3', Sox13: 5'-CTCCTCATGGACATCCATTTCATTC-3' and
403 control 5′-ATGGCCTCGGAGCTGGAGAGCCTCA-3′. Electroporation was performed in a chamber
404 using five 5V pulses of 50ms in 100ms intervals[45].

405

406 **Whole mount and hybridization chain reaction fluorescent *in situ* hybridisation**

407 *In situ* hybridization was carried out following previously described protocols[48]. Whole-mount pictures
408 were taken using an Olympus SZX12 with a Retiga2000R camera and Q-Capture Pro7 software.

409 HCR v3 was performed using the Molecular Technologies protocol[21]. Briefly, embryos were
410 fixed in 4% PFA for one hour at room temperature, dehydrated in a series of methanol in PBT and
411 stored overnight at -20°C. After rehydration and proteinase-K treatment (20mg/ml; 3 min) embryos
412 were post-fixed in 4% PFA for 20 min. Embryos were then washed on ice in PBS, 1:1 PBT/5X SSC
413 (5X sodium chloride sodium citrate, 0.1% Tween-20) and 5X SSC for 5 min each. Pre-hybridization
414 in hybridisation buffer was performed for 5 min on ice, followed by 30 minutes at 37°C. Embryos were
415 hybridised overnight at 37°C with probes at 4pmol/ml in hybridization buffer. After four 15 min washes
416 with probe wash buffer at 37°C, preamplification was carried out in amplification buffer for 5 min at
417 room temperature. Hairpins were prepared individually at 30pmol final concentration; they were
418 incubated at 95°C for 90 seconds followed by cooling to room temperature for 30 min, protected from
419 light. Cooled hairpins were added to 500µl amplification buffer, embryos were incubated in hairpins
420 overnight at room temperature followed by two 5 min and two 30 min washes in 5X SSC. After a 5
421 min incubation in DAPI (10 mg/ml), they were washed three times for 10 min with 5X SSC, before
422 being imaged using Leica SP5 laser scanning confocal inverted microscope using the LAS AF
423 software.

424 For HCR image analysis, Z-stacks were collected for 50-70µm; figures show projections of all
425 stacks. Images were processed using ImageJ and the ImageJ Plot Profile tool was used to calculate
426 intensity plots. In brief, an optical section in the centre of the placode territory was selected using the

427 Pax2 channel as a reference. The Pax2 channel was added to the Sox8, Lmx1a, Foxi3 or Tfap2e
428 channels, respectively. Intensity values were then calculated across the area of interest and plotted.
429

430 **Whole mount immunostaining**

431 Embryos were collected in PBS, fixed for 25 minutes at room temperature in 4% PFA, washed in
432 PBS-Tx (PBS + 0.2% Triton X-100) and blocked in 5% goat serum in PBS-Tx for 3–5 h at room
433 temperature. Embryos were then incubated in primary antibody diluted in blocking buffer for 24-72
434 hours at 4°C. Primary antibodies were rabbit anti–mCherry (1:200; Abcam ab167453), mouse anti–
435 NF (1:100; Thermo Fisher Scientific 13-0700), rabbit anti-GFP (1:500; Abcam a11122) or mouse anti
436 GFP (1:1000; Molecular Probes A11120). After five 60 min washes and one overnight wash in PBS-
437 Tx, embryos were incubated in secondary antibodies (1:800) at 4°C overnight. Secondary antibodies
438 used were goat anti–rabbit IgG Alexa Fluor 488 (Thermo Fisher Scientific, A11036), donkey anti–
439 rabbit Alexa Fluor 568 (Thermo Fisher Scientific, A11001), goat anti mouse IgG Alexa Fluor 488
440 (Molecular Probes A11001) and goat anti mouse IgG Alexa Fluor 568 (Molecular Probes A11004).
441 Embryos were then were briefly incubated in PBS containing 10 mg/ml DAPI and washed at least five
442 times in PBS-Tx before being mounted on slides and imaged using Leica SP5 laser scanning confocal
443 inverted microscope using a 10x objective or an Olympus SZX12 with a Retiga2000R camera and Q-
444 Capture Pro7 software. Confocal whole mount images in the manuscript are maximum intensity
445 projections of embryo z-stacks. Sections were imaged using a 63x oil immersion objective and
446 maximum intensity projections are shown.
447

448 **Cryosectioning**

449 Embryos were embedded in gelatine as previously described[49] and cryo-sectioned at 15-20μm using
450 a Bright OTF5000 cryostat. Sections were mounted using Mowiol 4-88 (Sigma Aldrich, 81381) and
451 imaged using Leica SP5 laser scanning confocal inverted microscope (LAS AF software) or a Zeiss
452 Axiovert 200M microscope with a Hamamatsu C4742-95 camera and using OCULAR software.
453

454 **Quantitative PCR (qPCR)**

455 RNA from dissected otic tissue was isolated using the RNAqueous-Micro Kit (Ambion, AM1931) and
456 reverse transcribed. Primers for target genes were designed with PrimerQuest (IDT). qPCR was
457 performed using Aria Mx Real-Time System (Agilent Technologies) with SYBR green master mix
458 (Roche, 64913850001). RT-qPCR for antisense oligonucleotide experiments was carried out with a
459 minimum of three biological replicates; two sample Welch's t-tests with Benjamini-Hochberg multiple
460 test correction were used to determine statistical significance between control and antisense
461 oligonucleotide ΔCt values. Relative expression $(2^{-\Delta\Delta Ct})$[50] was calculated using the NormqPCR
462 package in R. The geometric mean of Gapdh and Rplp1 expression, or Rplp1 expression alone, was
463 used to normalise gene expression.
464
465 **Bulk RNA sequencing**
466 To label otic and epibranchial cells embryos were electroporated with Lmx1aE1-EGFP and Sox3U3-
467 EGFP plasmids, respectively, and whole heads were used for cell collection. For overexpression
468 experiments embryos were electroporated with Sox8-mCherry+Lmx1aE1-EGFP or with pCAB-
469 mCherry+pCAB-EGFP, the endogenous otic placode and the trunk were removed before cell
470 dissociation. Cells were dissociated in FACSmax cell dissociation solution (Ambion, T200100)
471 containing papain (30U/mg, Sigma-Aldrich, 10108014001) for 20min at 37°C before being transferred
472 to Hanks Balanced Solution without calcium and magnesium (HBSS, Life Technologies, 14185045)
473 containing 5% heat-inactivated foetal-bovine-serum (FBS), rock inhibitor (10 μM, Stemcell
474 Technologies, Y-27632) and non-essential amino acids (Thermo Fisher Scientific, 11140035). Cells
475 were disaggregated by pipetting, sequentially filtered through 0.35μm and 0.20μm filters (Miltenyi
476 Biotech, 130-101-812). Pelleted cells were resuspended in 500μl HBSS and isolated by fluorescent
477 activated cell sorting (FACS) using a BD FACS-Aria Diva. 2000 cells per biological replicate were
478 collected, centrifuged at 200xg for 5 mins at 4°C, washed with PBS, and resuspended in lysis buffer.
479 RNA was extracted using Ambion RNAqueous Micro Total RNA isolation kit (AM1931, ThermoFisher
480 Scientific). RNA integrity was checked using Bioanalyser with Agilent RNA 6000 pico kit (Agilent
481 Technologies, 5067-1513); samples with RIN >7 were processed for library preparation. Sequencing
482 libraries were prepared using Nextera Sample low input kit (Illumina, 15028212) and sequenced using

483 75bp paired end reads on the Illumina Hiseq4000 platform. A minimum of three biological replicates
484 were used for analysis.
485
486 **Single-cell RNA extraction, library preparation and sequencing**
487 HH6-7 embryos were electroporated with Pax2E1-EGFP to label OEPs; cells were dissociated as
488 described above and 288 $EGFP^+$ cells from each ss8-9, ss11-12 and ss14-15 were collected by FACS
489 in 96-well plates. Sequencing libraries were prepared following the SmartSeq2 protocol[51]. Libraries
490 were sequenced on the Illumina NextSeq500 platform using single end 75bp sequencing (ss8-9) or
491 on the HiSeq4000 platform using paired end 75bp sequencing (ss11-12, ss14-15).
492
493 **Nuclei isolation, ATAC library preparation and sequencing**
494 The ATACseq library was prepared following published protocols[52]. Approximately 30 pieces of the
495 OEP territory were dissected from ss8-9 embryos, dissociated with Dounce homogenizer (tight pestle)
496 in lysis buffer (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% IGEPAL CA-630). Nuclei were
497 pelleted at 4000 rpm at 4°C for 10 min. After removing the lysis buffer, 1.25µl Tn5 transposase
498 (Illumina, FC-131-1024) were added to 25µl reaction volume and incubated at 37°C for 10 mins.
499 Tagmented DNA was then purified with Mini Elute PCR purification kit (Qiagen, 28004) followed by 9
500 cycles of PCR enrichment using NEB High fidelity PCR kit (NEB, M0541S). The quality of ATAC
501 libraries was assessed with Agilent Bioanalyzer with DNA High Sensitivity kit (Agilent Technologies,
502 5067-4627) and quantified with Kapa NGS library quantification kit (Kapa Biosystems, KK4824). The
503 libraries were sequenced with Illumina HiSeq 2000/2500 in 2x100 cycles.
504
505 **H3K27ac, H3K27me3 and H3K4me4 ChIP library preparation and sequencing**
506 Approximately 200 pieces of OEP ectoderm were dissected from ss8-9 embryos. The tissue was
507 dissociated in nuclei expulsion buffer (0.5% NP-40, 0.25% Triton X-100, 10 mM Tris-HCl, pH 7.5,
508 3mM $CaCl_2$, 0.25M sucrose, 1x protease inhibitor (Roche), 1mM DTT, and 0.2mM PMSF) using a
509 Dounce homogenizer (loose pestle); crosslinking was performed in 1% formaldehyde for 9 minutes,
510 followed by quenching with 1M glycine for 5 minutes. Cells were then pelleted and washed 3 times

511 with PBS containing protease inhibitor (Roche, 11873580001). The pellets were snap-frozen and
512 stored at -80°C for later use. Crosslinked chromatin was fragmented by sonication in an ice bath
513 (Misonix Q700 at 7 Amplitude, 5x40 seconds, 30 seconds on, 60 seconds off) and immunoprecipitated
514 following the Nano-ChIP protocol[53]. For chromatin immunoprecipitation (ChIP), the following
515 antibodies were used anti-H3K27Ac (Abcam, ab4729), anti-H3K4me3 (Diagenode, A5051-001P) and
516 anti-H3K27me3 (Millipore, 07449). Adaptors and primers from NEBNext library preparation kit
517 (Illumina, E6040S) were used to prepare the library following NanoChip protocol[53]. The library was
518 enriched using 14 PCR cycles and quantified with Kapa NGS library quantification kit (Kapa
519 Biosystems, KK4824) before pooling at a concentration of 20nM. Library quality was assessed with
520 Agilent Bioanalyzer with DNA High Sensitivity kit (Agilent Technologies, 5067-4627) and quantified
521 with Kapa NGS library quantification kit (Kapa Biosystems, KK4824). The libraries were sequenced
522 with Illumina HiSeq 2000/2500 in 2x100 cycles.
523
524 **High-Throughput-Sequencing-Data (HTSD) Analysis**
525 All data alignment and downstream analysis was carried out using NF-core and custom Nextflow
526 pipelines to allow full reproducibility. All code used, including Nextflow pipelines and downstream
527 analysis, can be found at https://github.com/alexthiery/otic-reprogramming. Full detailed
528 documentation for the pipeline is also available at https://alexthiery.github.io/otic-reprogramming/. A
529 custom Docker container used for the downstream analysis pipeline can be found at
530 https://hub.docker.com/repository/docker/alexthiery/otic-reprogramming-r_analysis. This also allows
531 for interactive exploration of the data.
532
533 **Bulk-RNA-seq**
534 Bulk-RNA-seq data were processed and aligned to GalGal6 using the default NF-core RNA-seq (v2.0)
535 pipeline[54] which uses the STAR aligner. Downstream differential expression analysis (Lmx1aE1-
536 EGFP vs Sox3U3-EGFP; Sox8OE vs ControlOE) was carried out with the DESeq2 package in R[55].
537 Adjusted p-values were calculated using the default DESeq2 multiple test correction (Benjamini-

538 Hochberg). Differentially expressed transcripts were determined by an absolute log2 fold-change >1.5
539 and adjusted p-value < 0.05.
540
541 **Single-cell-RNA-seq alignment**
542 SmartSeq2 single-cell-RNA-seq reads were aligned and processed using a custom Nextflow DSL2
543 pipeline. Adaptor sequences were trimmed using Cutadapt (v2.10). HISAT2 (v2.2.1) was then used
544 to build a genome index from GalGal6 (amended to include a GFP sequence), before extracting splice
545 sites from the GTF and aligning reads. Read counts were obtained using HTSeq (v0.12.4). BAM files
546 from HISAT2 were also passed to Velocyto (v0.17)[20] in order to get spliced and unspliced expression
547 matrices for further downstream analysis.
548
549 **Single-cell-RNA-seq data analysis**
550 Downstream data analysis was carried out primarily using the Antler R package (version:
551 Development2019)[18]. We excluded from the dataset: cells which expressed fewer than 1k genes or
552 fewer than 500k reads, cells with more than 6% reads from mitochondrial genes, genes which were
553 expressed in fewer than 3 cells and genes with CPM < 10.
554
555 *Identification of gene modules*
556 To identify clusters of genes with correlated expression unbiasedly (gene modules), we used the
557 Antler *identifyGeneModules* function. Genes which had a Spearman correlation greater than 0.3 with
558 at least 3 other genes were first removed. Genes were then hierarchically clustered iteratively into
559 gene modules and filtered. Gene modules were filtered based on the minimum expression level (5
560 CPM) and the proportion (0.4) of cells expressing a gene module. The number of final gene modules
561 was set to 40 to achieve reasonably large clusters, which broadly characterise cell type diversity
562 across the dataset. Gene modules were then filtered based on the presence of genes with known
563 expression profiles in ectodermal derivatives, before cells were re-clustered based on the remaining
564 gene modules. We used the expression of known markers to assign cell states. Placodal cells were

565 then subset and re-clustered using a new set of gene modules identified from the subset cell
566 population. Three main cell states were identified as otic, epibranchial and OEP populations.
567
568 *Gene expression dynamics at the otic-epibranchial branching point*
569 To model the transcriptional dynamics at the otic-epibranchial branching point, we ordered cells along
570 pseudotime using Monocle2[56]. The lineage tree was rooted to the earliest cell state (ss8-9) which was
571 comprised mostly of OEPs. The expression of genes along the bifurcation point was then modelled
572 using branch expression analysis modelling (BEAM).
573 To assess whether individual OEPs simultaneously express markers associated with the otic
574 or epibranchial state, we calculated the co-expression of key otic and epibranchial genes within each
575 of the Monocle branches. First, gene expression was binarized based on the presence or absence of
576 gene expression; then the proportion of cells co-expressing pairs of otic-epibranchial genes within
577 each of the branches was calculated. A Kruskal Wallis test was then used to compare the proportion
578 of cells co-expressing otic and epibranchial genes between the three branches. Post-hoc pairwise
579 comparisons between the OEP branch and the otic or epibranchial branch were carried out using two-
580 tailed Wilcoxon rank sum tests.
581 RNA velocity provides an alternative to Monocle2 for inferring future cell state of individual
582 cells. Spliced, un-spliced and spanning matrices obtained from Velocyto[20] were subset in R based on
583 genes and cells used to generate the Monocle trajectory. RNA velocity was then calculated using the
584 Velocyto.R package, where spanning reads were used to fit gene offsets. Single cell velocities as well
585 as vector fields were subsequently visualised on both tSNE and Monocle DDRTree embeddings.
586
587 **ChIP-seq and ATAC-seq alignment and peak calling**
588 ChIP-seq and ATAC-seq data were processed and aligned to GalGal6 using the default NF-core
589 ChIP-seq (v1.2.0) and NF-core ATAC-seq (v1.2.0) pipelines[54], respectively. Reads were aligned using
590 the Burrows-Wheeler aligner. MACS v2.2.7.1 was used to call broad peaks (FDR < 0.1) and narrow
591 peaks (FDR < 0.05) for ChIP-seq and ATAC-seq data, respectively.
592

593 **Enhancer discovery**

594 Putative enhancers were identified using a custom Nextflow DSL2 pipeline. First, bedtools (v2.29.2)
595 was used to subset ATAC-seq narrow peaks, which overlap with a broad H3K27ac peak, whilst
596 removing those overlapping with broad H3K27me3 peaks. Remaining ATAC-seq peaks were then
597 annotated with Homer (v4.11), using a GalGal6 GTF which was filtered for protein coding genes.
598 Promoter and exonic peaks were then removed. Motif enrichment and functional enrichment analysis
599 were carried out using Homer and g:Profiler, respectively.

600

601 **Data availability**

602 All data have been deposited in Gene Expression Omnibus (GEO; Accession number GSE168089).
603 Nextflow pipelines and downstream analysis, can be found at https://github.com/alexthiery/otic-
604 reprogramming. Full detailed documentation is available at https://alexthiery.github.io/otic-
605 reprogramming/. A custom Docker container used can be found at
606 https://hub.docker.com/repository/docker/alexthiery/otic-reprogramming-r_analysis.

607

608 **References**

609

610 44 Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick
611 embryo. *J Morphol* **88**, 49-92 (1951).

612 45 Norris, A. & Streit, A. Morpholinos: Studying gene function in the chick. *Methods* **66**, 454-
613 465, doi:10.1016/j.ymeth.2013.10.009 (2014).

614 46 Chapman, S. C., Collignon, J., Schoenwolf, G. C. & Lumsden, A. Improved method for chick
615 whole-embryo culture using a filter paper carrier. *Dev Dyn* **220**, 284-289, doi:10.1002/1097-
616 0177(20010301)220:3<284::AID-DVDY1102>3.0.CO;2-5 (2001).

617 47 Nagai, H., Lin, M. C. & Sheng, G. A modified cornish pasty method for ex ovo culture of the
618 chick embryo. *Genesis* **49**, 46-52, doi:10.1002/dvg.20690 (2011).

619 48 Streit, A. & Stern, C. D. Combined whole-mount in situ hybridization and
620 immunohistochemistry in avian embryos. *Methods* **23**, 339-344,
621 doi:10.1006/meth.2000.1146 (2001).
622 49 Stern, C. D. & Holland, P. W. H. *Essential Developmental Biology: A Practical Approach.*,
623 (1993).
624 50 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time
625 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408,
626 doi:10.1006/meth.2001.1262 (2001).
627 51 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-
628 181, doi:10.1038/nprot.2014.006 (2014).
629 52 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition
630 of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
631 binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218,
632 doi:10.1038/nmeth.2688 (2013).
633 53 Adli, M. & Bernstein, B. E. Whole-genome chromatin profiling from limited numbers of cells
634 using nano-ChIP-seq. *Nat Protoc* **6**, 1656-1668, doi:10.1038/nprot.2011.402 (2011).
635 54 Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines.
636 *Nat Biotechnol* **38**, 276-278, doi:10.1038/s41587-020-0439-x (2020).
637 55 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
638 RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
639 56 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat*
640 *Methods* **14**, 979-982, doi:10.1038/nmeth.4402 (2017).

641

642 **Acknowledgements.** We thank Tatjana Sauka-Spengler and Ruth Williams for assistance with
643 establishing ChIPseq and scRNAseq protocols, Monica Tambalo for help with tissue collection,
644 Rosalinda Guerra and Ewa Kolano-Merlin for technical assistance, and James Briscoe for advice on
645 single cell transcriptomics. Thanks to Claudio D. Stern for comments on the manuscript and the

Streit group for discussions. This work was funded by grants to AS from the Biotechnology and Biological Sciences Research Council (BB/S005536/1; BB/M006964/1) and the National Institute on Deafness and Other Communication Disorders (DC011577).

Author contributions. A.L.B., J.C. and A.S. designed experiments; A.L.B. performed RNAseq for Lmx1a-E1-EGFP, Sox3U3-EGFP, Sox8-mCherry/Lmx1aE1-EGFP, mCherry/EGFP populations, scRNAseq, HCR analysis, Sox8/Pax2/Lmx1a OE and some Sox8 OE experiments; J.C. performed ATACseq, ChIPseq, enhancer cloning and testing, in situ hybridisation, qPCR, knock-down and Sox8 OE experiments; J.C. performed the initial analysis of ATACseq and ChIPseq data. A.L.B. interpreted and supervised bioinformatics analysis; J.D. performed some scRNAseq analysis, A.T. analysed all transcriptomic and epigenomic data and established the integrated pipeline; A.L.B. generated all figures; A.L.B., A.T. and A.S. wrote the manuscript.

Competing interests. The authors declare no competing interests.

Additional information

Supplementary information is available for this paper.

Correspondence and requests for material should be addressed to Andrea Streit: andrea.streit@kcl.ac.uk.